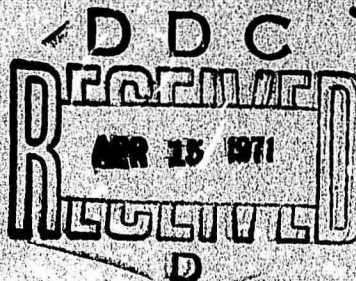


AD721459

**HUMAN PERFORMANCE CENTER  
DEPARTMENT OF PSYCHOLOGY****The University of Michigan, Ann Arbor****Theoretical Entities  
in Statistical Explanation****JAMES G. GIBSON**

1. This document has been approved for public release and sale; its distribution is unlimited.

Reproduced by  
**NATIONAL TECHNICAL  
INFORMATION SERVICE**  
Springfield, Va. 22151

**Memorandum Report No. 12****October 1970**

**BEST  
AVAILABLE COPY**

## THE HUMAN PERFORMANCE CENTER

### DEPARTMENT OF PSYCHOLOGY

The Human Performance Center is a federation of research programs whose emphasis is on man as a processor of information. Topics under study include perception, attention, verbal learning and behavior, short- and long-term memory, choice and decision processes, and learning and performance in simple and complex skills. The integrating concept is the quantitative description, and theory, of man's performance capabilities and limitations and the ways in which these may be modified by learning, by instruction, and by task design.

The Center issues two series of reports. A Technical Report series includes original reports of experimental or theoretical studies, and integrative reviews of the scientific literature. A Memorandum Report series includes printed versions of papers presented orally at scientific or professional meetings or symposia, methodological notes and documentary materials, apparatus notes, and exploratory studies.

ACCESSION: 27		
CSSTI	WRITE SECTION	<input checked="" type="checkbox"/>
DDCC	BUFF SECTION	<input type="checkbox"/>
PLAN. (REQ.)		<input type="checkbox"/>
JUSTIFICATION.....		
.....		
DISTRIBUTION/AVAILABILITY CODES		
DIST.	AVAIL	and/or SPECIAL
A		

THE UNIVERSITY OF MICHIGAN  
COLLEGE OF LITERATURE, SCIENCE AND THE ARTS  
DEPARTMENT OF PSYCHOLOGY

THEORETICAL ENTITIES IN STATISTICAL EXPLANATION

James G. Greeno

HUMAN PERFORMANCE CENTER--MEMORANDUM REPORT NO. 12

OCTOBER 1970

A talk based on this paper was given at meetings of the Philosophy of Science Association, Boston, October, 1970. The research was supported by the Advanced Research Projects Agency, Department of Defense, monitored by the Air Force Office of Scientific Research, under Contract No. AF 49(638)-1736 with the Human Performance Center, Department of Psychology, University of Michigan.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. This document has been approved for public release and its distribution is unlimited.

The ideas in this paper are based on an analysis of statistical explanation that uses the information transmitted by a theory [2]. Consider a theory that specifies a probability distribution on . . . events of some domain, where for purposes of analysis we divide the variables that describe the domain into two sets: [M], a set of variables whose values are to be explained, and [S], a set of variables whose values are used to explain the values of the variables in [M]. The information transmitted by the theory is

$$I(S,M) = H(S) + H(M) - H(S \times M), \quad (1)$$

where  $H(S)$  and  $H(M)$  are the uncertainties of the events in [S] and [M], respectively, and  $H(S \times M)$  is the uncertainty of the joint events.

To illustrate this idea, consider the problem of explaining medical symptoms such as fever, coughing, skin rash and abdominal pain. The kinds of variables available for explaining symptoms are facts in a patient's medical history and information about the patient's recent contact with other people with similar symptoms. For the moment, I will avoid reference to disease entities, and deal only with the data that are available to the physician. I have to strain your imagination a bit to do this, but a theory about skin rash and fever might run something like this: Let  $M_1$  denote a combination of fever and a blotchy skin rash, and let  $M_2$  denote the absence of this pattern of symptoms. Let [S] be a set of three variables: (1) the patient's age, (2) whether he has been in contact with another person showing the symptoms within the last month, and (3) whether he had the symptoms himself at any previous time. If each of these variables had three values, then [S] would partition the domain of people into  $3^3 = 27$

sets: for example, one set could include all infants under the age of two months who had been in contact with someone with the symptoms recently and had not shown the symptoms themselves. The theory then would consist of a set of probabilities of the 27 events distinguished in the explanans, and 27 conditional probabilities--a value of  $P(M_1|S_i)$  for each  $S_i \in [S]$ . These probabilities are sufficient to specify the probabilities of all the joint events, and therefore the overall probability of  $M_1$  is specified also. Facts like "infants less than two months old seldom have the symptoms" and "contact with a person who has the symptoms increases the probability of having them, unless the person has had the symptoms previously himself" would be incorporated in the conditional probabilities, and facts like the proportion of people who have had the symptoms and the proportion of people who are younger than two months of age would be incorporated in the probabilities of the explanans.

Let

$$a_i = P(S_i), c_k = P(M_k), p_{ik} = P(M_k|S_i). \quad (2)$$

In relation to the example about symptoms, the  $a_i$  are the proportions of people in the various categories specified by age, medical history, and recent contacts. The  $c_k$  are the proportions of people who have or do not have the symptoms now. The  $p_{ik}$  are the conditional probabilities of having the symptoms, given the various categories of age, medical history, and recent contacts. The quantities in equation (1) are defined as

$$\begin{aligned} H(S) &= \sum_i -a_i \log a_i, \quad H(M) = \sum_k -c_k \log c_k, \\ H(S \times M) &= \sum_{ik} -a_i p_{ik} \log a_i p_{ik}. \end{aligned} \quad (3)$$

If we think about a theory in relation to its information transmitted, we are immediately led to considering the overall properties of the theory. Information transmitted is a measure of the reduction in uncertainty brought about by the dependencies between the variables. Thus, it is an index of the explanatory power of the theory in relation to the entire domain of events that the theory deals with. In fact, the point of introducing the analysis based on information transmitted was to provide some concepts that would make it reasonable to consider the evaluation of theories, rather than of single explanations. In my opinion, it is more appropriate and useful to consider general properties of theories than it is to deal with the status of single explanations. And the information transmitted seems to capture some of the properties that are desirable for a measure that is used to evaluate a theory. For example, a higher value of information transmitted generally goes with a greater degree of testability in Popper's sense, and a greater degree of predictive usefulness.

In this paper, I want to extend this line of analysis to the consideration of theories that provide statistical explanations and that postulate theoretical entities. (The earlier analysis was limited to relationships between empirical variables.) To carry out this analysis, I need to introduce a structure that is slightly more complex than the one described above.

Consider a domain  $\Omega$ , partitioned by three sets of variables:  $[S]$ , a set of empirical variables used as explanans;  $[M]$ , a set of empirical variables whose values are to be explained; and  $[T]$ , a set of postulated theoretical variables. The values of the theoretical variables are assumed to be produced by the values of the explanans, and in turn to

produce the values of the explananda, according to statistical laws specified by the theory. The sense in which I use the phrase "theoretical variables produce the values of the explananda" is entirely neutral as regards metaphysics. I simply mean that the conditional probabilities of the empirical outcomes--the explananda--given any theoretical state are the same, regardless of the value of the explanans that applies. In other words, I assume that the probability law connecting any given theoretical state with the explananda is independent of the conditions that produced that theoretical state. Under this assumption, the theory consists of three sets of probabilities: (1) a vector of probabilities  $a_i$ , where

$$a_i = P(S_i), \quad (4)$$

(2) a set of conditional probabilities  $q_{ij}$ , linking the explanans to the theoretical states, where

$$q_{ij} = P(T_j | S_i), \quad (5)$$

and (3) a set of conditional probabilities  $r_{jk}$ , linking the theoretical states to the explananda, where

$$r_{jk} = P(M_k | T_j). \quad (6)$$

First, it may be remarked that these quantities relate to those described at the beginning in a straightforward way. When theoretical variables are not taken into account, a theory is specified by a vector of probabilities of the explanans, the  $a_i$ , and a matrix of conditional probabilities linking the explanans and the explananda, the  $p_{ik}$ . This matrix is just the product of the matrices of conditional probabilities  $q_{ij}$  and  $r_{jk}$ . In

other words, the probabilities linking the explanans and the explananda are specified by the probabilities defined in equations (5) and (6).

$$P_{ik} = \sum_{j=1}^t q_{ij} r_{jk}. \quad (7)$$

Of course, a theory is testable because data can be used to check its assumptions. In a theory of the kind described here, the data are the empirical values of the  $a_i$  and the  $p_{ik}$ , and these can be used to test the theory. If the theory specifies numerical probabilities (rather than merely the existence of specified states) then the process of testing the theory is just that of comparing the theoretical values of the  $a_i$  and the values of  $p_{ik}$  calculated from equation (7) with empirical values of  $a_i$  and  $p_{ik}$  that can be obtained by whatever means are available. When the theory just specifies that certain states exist, testability involves the estimation of parameters, and I will leave that discussion for a later time in this paper.

The main question that I want to deal with is the relationship between theoretical variables and the information-theoretical properties of a theory. The situation involves three matrices: one linking the empirical explanans and the theoretical variables, a second involving the theoretical variables and the empirical explananda, and a third, the product of the first two, linking the empirical explanans and the empirical explananda. We can calculate the information transmitted by each of these. As an hypothetical example, consider a theory that specifies four possible values of  $S$  and three possible values of  $M$ . This might, for example, involve classification of medical information into four categories of medical history and three categories of symptom patterns. In addition, the theory postulates the existence of a theoretical state, such as a

disease entity that may be present or absent, giving two values of  $T$ . To have a concrete illustration, suppose the values of the  $a_i$ ,  $q_{ij}$ , and  $r_{jk}$  are

$$[a_i] = [.20, .30, .10, .40],$$

$$[q_{ij}] = \begin{array}{c|cc} & T_1 & T_2 \\ \hline S_1 & .70 & .30 \\ S_2 & .60 & .40 \\ S_3 & .30 & .70 \\ S_4 & .20 & .80 \end{array} \quad [r_{jk}] = \begin{array}{c|ccc} & M_1 & M_2 & M_3 \\ \hline T_1 & .70 & .20 & .10 \\ T_2 & .20 & .20 & .60 \end{array} \quad (8)$$

The values of information transmitted by these matrices, using natural logarithms, are

$$I(S,T) = 0.10, \quad I(T,M) = 0.16.$$

The implications of this theory for dependencies between explanans and explananda are described by the values of  $p_{ik}$ , which are

$$[p_{ik}] = \begin{array}{c|ccc} & M_1 & M_2 & M_3 \\ \hline S_1 & .55 & .20 & .25 \\ S_2 & .50 & .20 & .30 \\ S_3 & .35 & .20 & .45 \\ S_4 & .30 & .20 & .50 \end{array} \quad (9)$$

Using the values of  $a_i$  given earlier, the information transmitted is

$$I(S,M) = 0.03.$$

An interesting general fact is illustrated by the example. The value

of  $I(S,M)$  is never larger than either  $I(S,T)$  or  $I(T,M)$ . This claim can be proved as follows. Consider a state-space  $[Q] = [S] \times [T]$ . The conditional probabilities  $P(M_k|Q_h)$  will equal the conditional probabilities  $r_{jk}$ , hence, the information transmitted  $I(Q,M)$  will be the same as  $I(T,M)$ . (This is established by an argument given in [2] in connection with equation (17) of that paper.) The state-space  $[S]$  is a partition of  $[Q]$ , and it can be shown that the information transmitted by collapsing two or more of the states of a matrix cannot be larger than the information transmitted by the original matrix.

What this fact means is that the dependencies between the theoretical variables and the empirical variables are stronger in the sense of approaching probabilities of one and zero than are the dependencies between the two sets of empirical variables. I am inclined to believe that this fact is related to the intuitions that we have regarding the desirability of theoretical explanations when the dependencies between empirical variables are statistical. If the theory correctly specifies a set of states that produce the phenomena to be explained in the sense assumed here, then the explanation in terms of the theoretical states is better, in the sense of information transmitted, than is the explanation in terms of the empirical explanans.

On the other hand, the improved quality of the explanation obtained by introducing theoretical states may be merely a "paper" profit. Without further theoretical development, merely introducing theoretical variables may not change the empirical content of the theory. However, further developments are often guided by the postulated properties of theoretical entities.

One kind of development involves the discovery of new empirical variables

that can be added to the explanans. Refer back to equation (8), and imagine that at some stage of scientific investigation the probabilities specified there represent the best available theory about the explananda [M]. This is equivalent to saying that the best judgment that can be made on the available evidence is that there are states  $T_1$  and  $T_2$  that cannot be distinguished directly in observations (at least with present technology) that are related to the explananda according to probabilities given by the values of the  $r_{jk}$ . One research problem that would be potentially worthwhile in such a situation would be the search for improved knowledge about the conditions that produce the theoretical states. If additional variables that are related to the theoretical states could be discovered, then the likely outcome would be an increase in  $I(S,T)$ , and a corresponding increase in  $I(S,M)$ . The limit of this process of obtaining better knowledge about conditions producing the theoretical states is a theory with information transmitted equal to  $I(T,M)$ .

A second kind of research that could be motivated by a state of knowledge such as that described by equation (8) involves an effort to develop new theoretical variables in order to increase  $I(T,M)$ . If the conditional probabilities of the explananda, given the theoretical states are substantially different from one and zero, there is a strong presumption that the theoretical description is incomplete and that additional distinctions among theoretical states can be found to reduce the conditional uncertainty of the explananda given the theoretical states.

A third kind of development from a situation like that of equation (8) could be the discovery of new phenomena that can be explained by the empirical explanans and the theoretical variables of the theory. The extension of a theory to additional explananda usually has the effect of increasing the information transmitted--in this case, involving the relationship between theoretical states and the explananda.

Next, I will discuss theories of the form of equation (8), but with free parameters rather than numerical values of the conditional probabilities  $q_{ij}$  and  $r_{jk}$ . Theoretical proposals that specify numerical probabilities are very rare. However, a kind of situation that occurs frequently is one in which dependencies among empirical variables are known, and a theorist proposes to explain the dependencies in relation to a set of theoretical states. In its weakest form, this kind of theoretical proposal simply specifies a number of theoretical states and all the conditional probabilities are free parameters. Let  $n$  be the number of values taken by the explanans and let  $m$  be the number of values taken by the explananda. Then if  $t$  is the number of postulated states in the theory, the form of the hypothesis is

$$\begin{aligned} \exists(q_{ij}: i=1, \dots, n; j=1, \dots, t) \quad \exists(r_{jk}: j=1, \dots, t; k=1, \dots, m) \\ (\forall i)(\forall k)[p_{ik} = \sum_{j=1}^t q_{ij} r_{jk}]. \end{aligned} \tag{10}$$

That is, the theory asserts that it is possible to find a set of probabilities (the  $q_{ij}$  and  $r_{jk}$ ) such that all the values of  $p_{ik}$  can be calculated using equation (7).

An hypothesis of this kind can be tested if the number of free parameters is smaller than the number of quantities that can be obtained from

the empirical dependencies. In the weak form of the theory described above, the number of free parameters is  $(t-1)n+(m-1)t$ . The number of empirical quantities is  $(m-1)n$ . Then the theory is testable if the following inequality is satisfied:

$$t < \frac{mn}{m+n-1} . \quad (11)$$

It should be noted that equation (11) applies only when a theory is stated without constraints on the theoretical parameters. Most frequently, substantive hypotheses about the postulated states impose constraints that reduce the number of free parameters of the theory. The most general form of the kind of theoretical proposal we are discussing specifies a set of free parameters  $(e_1, \dots, e_s)$  such that each of the conditional probabilities of the theory is a specified function of the parameters. Then the condition for testability is that  $s$  must be less than the number of empirical quantities, and the hypothesis has the form

$$\exists e_1) \dots \exists e_s) (\forall i) (\forall k) [p_{ik} = \sum_{j=1}^t q_{ij}(e_1 \dots e_s) r_{jk}(e_1 \dots e_s)] . \quad (12)$$

The procedure for testing the hypothesis involves finding estimates of the theoretical parameters that bring the values of  $p_{ik}$  implied by the theory as close as possible to the empirical values. Then the degree of approximation between the theoretical and empirical values of  $p_{ik}$  can be evaluated using standard statistical techniques.

There is enough structure in these general concepts now so that a realistic example can be introduced. I will describe a theory about the detection of weak signals proposed by Luce [4]. The theory is used in analysis of experiments where on some trials a relatively weak stimulus signal with some fixed energy level is presented along with a noisy input

that makes it difficult to tell whether the signal is there or not. The noise has a fixed mean energy level, and on trials when the signal is not presented the noise is presented alone. There are several conditions designed to produce different response biases. These may be produced by varying the payoffs for different kinds of correct responses (identifying signals when they are present or correctly saying that the signal is absent) or by imposing varying penalties for different kinds of errors (missing signals or saying there is a signal when only noise is presented). Another way of producing response bias is to vary the overall proportion of trials when a signal is presented, thus producing higher or lower expectations of the signal. The empirical explananda are the subjects' responses: on each trial a subject says either "yes" or "no," depending on whether he judges a signal to have been present or absent. The explanans are the experimental conditions: on each trial, there is some condition of payoff and a priori expectation of a signal, and either the signal is presented or only noise is presented. At the level of data, an experiment can be described as a set of conditional frequencies

$$p_{ik} = P(\text{yes} | \text{condition } i)$$

where the conditions are given in the following equation:

$$[p_{ik}] = \begin{array}{cc} & \begin{array}{cc} \text{yes} & \text{no} \end{array} \\ \begin{array}{c} S_1 \\ SN_1 \\ S_2 \\ SN_2 \\ \vdots \\ S_g \\ SN_g \end{array} & \begin{array}{cc} \begin{array}{c} P_{11} \\ P_{21} \\ P_{31} \\ P_{41} \\ \vdots \\ P_{(2g-1)1} \\ P_{(2g)1} \end{array} & \begin{array}{c} P_{12} \\ P_{22} \\ P_{32} \\ P_{42} \\ \vdots \\ P_{(2g-1)2} \\ P_{(2g)2} \end{array} \end{array} \end{array} \quad (13)$$

where  $g$  is the number of different motivational conditions.

Luce's theory uses the assumption of a threshold of detection and the probability of exceeding the threshold depends only on whether the signal was presented or not. The other theoretical variable is determined by the motivational conditions. Thus, on each trial, the subject is assumed to be in one of  $2g$  theoretical states, as specified in equation (14).

$[q_{ij}] =$ 

	$\bar{D}_1$	$D_1$	$\bar{D}_2$	$D_2$	.	.	.	$\bar{D}_g$	$D_g$
$S_1$	$1-q$	$q$	$0$	$0$	.	.	.	$0$	$0$
$SN_1$	$1-p$	$p$	$0$	$0$	.	.	.	$0$	$0$
$S_2$	$0$	$0$	$1-q$	$q$	.	.	.	$0$	$0$
$SN_2$	$0$	$0$	$1-p$	$p$	.	.	.	$0$	$0$
.	.	.	.	.				.	.
.	.	.	.	.				.	.
.	.	.	.	.				.	.
$S_g$	$0$	$0$	$0$	$0$	.	.	.	$1-q$	$q$
$SN_g$	$0$	$0$	$0$	$0$	.	.	.	$1-p$	$p$

(14)

$\bar{D}_h$  denotes a state in which the threshold of detection is not exceeded in motivational condition  $h$ , and  $D_h$  denotes a state in which the threshold is exceeded in motivational condition  $h$ . The probability of exceeding the threshold when the signal is presented is  $p$ , and the probability of exceeding the threshold when noise is presented alone is  $q$ .

It is assumed that the motivational states affect the relationships between the theoretical states and the responses. The states are assumed to be ordered in their tendency to produce biases favoring "yes" responses; let State 1 denote the condition producing the greatest reluctance to say "yes." It is assumed that there is some motivational state  $f$  such that

$$\begin{aligned} P(\text{yes}|\bar{D}_h) &= \begin{cases} 0 & \text{for } h \leq f \\ t_h & \text{for } h > f \end{cases} \\ P(\text{yes}|D_h) &= \begin{cases} u_h & \text{for } h \leq f \\ 1 & \text{for } h > f \end{cases} \end{aligned} \tag{15}$$

with  $u_h$  and  $t_h$  ordered monotonically with the values of  $h$ . In other words, for states in which the subject is reluctant to say "yes," he always says "no" if the threshold is not exceeded and divides his responses between "yes" and "no" randomly when the threshold is exceeded. And for states in which the subject is reluctant to say "no" he always says "yes" if the threshold is exceeded and divides his responses between "yes" and "no" when the threshold is not exceeded. The matrix of conditional probabilities connecting the theoretical states and the responses is given below.

$$[r_{jk}] = \begin{array}{cc|cc} & & \text{yes} & \text{no} \\ \hline \bar{D}_1 & & 0 & 1 \\ D_1 & & u_1 & 1-u_1 \\ . & & . & . \\ . & & . & . \\ . & & . & . \\ \bar{D}_f & & 0 & 1 \\ D_f & & u_f & 1-u_f \\ \bar{D}_{f+1} & & t_{f+1} & 1-t_{f+1} \\ D_{f+1} & & 1 & 0 \\ . & & . & . \\ . & & . & . \\ . & & . & . \\ \bar{D}_g & & t_g & 1-t_g \\ D_g & & 1 & 0 \end{array} \tag{16}$$

Recall that the theoretical values of the  $p_{ik}$  that can be compared with data are obtained by multiplying the matrices  $[q_{ij}]$  and  $[r_{jk}]$ . In

the case of Luce's theory, this yields

		yes	no
$[p_{ik}] =$	$S_1$	$qu_1$	$1-qu_1$
	$SN_1$	$pu_1$	$1-pu_1$
	.	.	.
	.	.	.
	.	.	.
	$S_f$	$qu_f$	$1-qu_f$
	$SN_f$	$pu_f$	$1-pu_f$
	$S_{f+1}$	$(1-q)t_{f+1}+q$	$(1-q)(1-t_{f+1})$
	$SN_{f+1}$	$(1-p)t_{f+1}+p$	$(1-p)(1-t_{f+1})$
	.	.	.
	.	.	.
	.	.	.
	$S_g$	$(1-q)t_g+q$	$(1-q)(1-t_g)$
	$SN_g$	$(1-p)t_g+p$	$(1-p)(1-t_g)$

(17)

Luce's theory illustrates the kind of situation described in connection with equation (12). A number of parameters are specified--there are  $g+2$  of them--and the theory asserts that the parameters determine the relationships between explanans (in this case, experimental conditions) and theoretical states and explananda (in this case, judgments about whether a signal was present). The parameters therefore specify a theoretical relationship between the explanans and the explananda. In order to obtain a numerical relationship, the parameters must be estimated from data, and the theory is testable if the number of free parameters is less than the number of empirical quantities in the data. In the present example, the number of empirical quantities is  $2g$ , so the theory is testable whenever  $g$  is greater than two.

Numerical values of the parameters must be estimated both to test

the theory and to specify the information-theoretical properties of the theory. Putting this in another way, the information transmitted by the theory is a function of the parameter values. To provide a further illustration of the application of information theory to the analysis of statistical explanations, I have calculated the information transmitted by one special case of the theory. Suppose that  $g = 5$  (that is, five different motivational conditions are used) and the five conditions are used equally often. Furthermore, suppose that signals are presented on one-half of the trials under each motivational condition. This means that each of the states constituting the explanans has probability .10, and from this we can calculate

$$H(S) = 2.30.$$

Now, suppose that an experiment is conducted and the estimated values of  $p$  and  $q$  are .60 and .30, respectively. Recall that these are the probabilities of exceeding the threshold on signal trials and noise trials, respectively. This permits us to calculate the probabilities of the theoretical states; the probability of each  $\bar{D}_h$  is .11 and the probability of each  $D_h$  is .09. We can then calculate

$$H(T) = 2.30, \quad H(S \times T) = 2.94.$$

The remaining parameters are the conditional probabilities of saying "yes" in the various theoretical states. Suppose it is estimated that  $f = 3$ , and the values of the  $u_h$  and  $t_h$  parameters are

$$u_1 = .40, \quad u_2 = .70, \quad u_3 = 1.00, \quad t_4 = .50, \quad t_5 = .80.$$

It turns out that this implies that subjects will say "yes" with probability

.51, and we can calculate

$$H(M) = .69, H(T \times M) = 2.55.$$

Combining these values using equations (1) and (7), we arrive at values of information transmitted as follows:

$$I(S,T) = 1.66, I(T,M) = 0.44, I(S,M) = 0.17.$$

Note again that the information transmitted by the relationships between explanans and explananda is smaller than either of the quantities involving the theoretical states.

Developments motivated by Luce's theory can be used to illustrate the remarks made earlier about the kinds of research that provide improvements in theories of this kind. One kind of development involves applying the theory to more complex experiments. Luce's theory has been applied to studies in which two different signals have been presented on different trials. The experiments involved auditory detection, and the signals were tones of different frequency. This change increases the number of states in the set of explanans, and thus increases the information transmitted by the theory. In addition, the subjects were asked to identify the signal after they judged whether a signal was presented. Thus, instead of just saying "yes" or "no," subjects also classified the stimuli as "high" or "low" as well. With a more detailed set of explananda, additional explanatory power was obtained. Of course, this experiment was not just cooked up to provide more detail for its own sake; the threshold theory makes the strong prediction that when the threshold is not exceeded (as is the case for all "no" responses when  $h > f$ ) the subject should have no information about which stimulus was

presented, and the experiment was designed in part to test this prediction.

The other kind of development involves adding to the complexity of the theoretical description, guided by data that are not consistent with the simpler theory. Krantz [3] has proposed such a theory, in which an additional state  $D_h^*$  is postulated for each motivational state.  $D^*$  is a state of strong detection, in which the subject is sure that a signal was presented. In Krantz' theory, then, there are two postulated thresholds. If the lower threshold is not exceeded the subject is in State  $\bar{D}_h$ , and if the higher threshold is exceeded the subject is in State  $D_h^*$ . It is assumed that the subject always judges that a signal was presented when he is in State  $D_h^*$ , regardless of the motivational state. The main reason for complicating a theory, of course, is to correct apparent defects that are revealed by failure of data to agree with predictions derived from the theory. But this also has the effect of increasing the explanatory power of the theory in the sense of information transmitted reflected by values of  $I(S,T)$  and  $I(T,M)$ .

This example from the theory of perception illustrates several aspects of the role of theoretical entities in statistical explanation. Other examples could be used for the same purpose, from other areas of psychological theory as well as from other fields. It also may be remarked that the analysis of discrete states does not affect the main features of the analysis. Systems that involve continuous variables can be analyzed from this point of view and their general properties are analogous to those of discrete-state systems, although the analysis of systems with discrete states is easier to understand.

An important feature of the kind of theory that I have been discussing

is that it describes a system whose statistical properties do not change over time. In such a stationary system, the main role of theoretical entities seems to be heuristic, in that they guide the development of new empirical and theoretical research and thus facilitate the extension of knowledge. In the remainder of this paper I will discuss systems that are not stationary, and in these systems the use of theoretical entities can lead to a considerable simplification of the statistical structure of a theory in addition to their heuristic value.

The simplest kind of nonstationary system is illustrated by experiments in learning and problem solving. In their simplest form, these experiments consist of repeated trials where a subject is given opportunities to study the material to be learned or is given information relevant to solving the problem. I will impose a relatively stringent condition of uniformity for the purpose of analysis here. I will ignore differences in procedure that occur on different trials. When such differences cannot be neglected, the situation would be analyzed as the concatenation of different experiments.

At the beginning of an experiment, the subject either gives only incorrect responses or he responds correctly with some probability due to guessing, depending on the procedure of the experiment. As the subject proceeds through the experiment, the probability of correct response increases. The subject may reach a state in which he gives only correct responses, or he may reach some asymptotic state in which the probability of correct response is at some level less than one. Thus, the most salient general feature of the experiment is an increase in the probability of correct response from some initial value  $c_1$  (where  $c_1$  may be zero) to some asymptotic level  $c_\infty$  (where  $c_\infty$  may be one).

The data from a learning or problem solving experiment are sequences

of responses given by the subjects. In general, the probability of response on each trial  $n$  depends on the sequence of responses given on the preceding  $n-1$  trials. One natural way to consider the experiment, then, is as a sequence of probabilistic events in which the response on each trial  $n$  is an event to be explained, and the sequence of responses on trials  $1, 2, \dots, n-1$  is the event available to be used as the explanans.

To illustrate the situation, consider the first four trials of an experiment. The data are hypothetical, and were generated from a theory that will be presented later. In the notation, a correct response is denoted 0 and an error is denoted 1. On trial 1, the subject either gives a correct response or an error, and this provides the first state-space for the explanans. The conditional probabilities  $p_{ik,2}$  are just the probabilities of correct response on trial 2. In the following equation, the numbers in parentheses to the left of the first-trial states are the probabilities of those states, and the probabilities of response on trial 2 are in parentheses at the top.

$$[p_{ik,2}] = \begin{array}{cc} & \begin{array}{cc} (.40) & (.60) \\ 0 & 1 \end{array} \\ \begin{array}{cc} (.25) & 0 \\ (.75) & 1 \end{array} & \left| \begin{array}{cc} .40 & .60 \\ .40 & .60 \end{array} \right. \end{array} \quad (18)$$

The information-theoretical quantities are

$$H_2(S) = 0.56, \quad H_2(M) = 0.67, \quad I_2(S,M) = 0.00.$$

The reason that no information is transmitted is that the response on trial 2 is independent of the response on trial 1 in this situation.

The responses on trial 3 are related to the sequences of response on trials 1 and 2.

$[p_{ik,3}] =$

the system eventually reaches a stable asymptotic level of response probability, the responses will eventually become independent of preceding sequences. In other words,

$$\lim_{n \rightarrow \infty} I_n(S, M) = 0.$$

This fact makes it reasonable to think about a theory of learning in relation to the sum of the values of  $I_n$  across trials. Using this line of thinking, our state of knowledge about a learning system would be evaluated in regard to the extent to which the performance of a learning subject could be predicted on the basis of his earlier performance, and this seems like a reasonable way to proceed. For example, it fits with our intuitions about discoveries that in fact count as additions to our knowledge. When responses are measured in more detail, as by a finer classification of errors or by measuring additional properties such as time to respond, the effect is to increase the values of  $I_n$ , for the same reason that additional variables added to the explanans and the explananda always increase information transmitted. In addition, when we analyze properties of the sequence of study trials or the information that is presented, thus going beyond the assumption of uniform trials that I have imposed on this analysis, we add new variables to the explanans and thus also increase the values of  $I_n$ .

These remarks about learning systems provide a framework for the methodological analysis of nonstationary systems that is consistent with the information-theoretical analysis worked out earlier for stationary systems. My remaining discussion will consider the role of theoretical entities in such systems.

The general structure that I will use for my remaining remarks involves

a state-space of postulated variables. On each trial there is a set of postulated states  $[T_n]$  related to the explanans  $[S_{n-1}]$  and the explananda  $[M_n]$  by specified probability laws given as matrices of conditional probabilities  $[q_{ij,n}]$  and  $[r_{jk,n}]$ . Thus on each trial the same kind of structure that we used earlier for stationary systems can be applied to analyze the information-theoretical properties of a non-stationary system.

In principle, the theoretical states may be as complex as the theorist wishes. However, a number of simplifying restrictions are frequently used. The first is that the theoretical state-space  $[T_n]$  is a constant that I will denote  $[T]$ . Secondly, the conditional probabilities of responses given theoretical states are constant, so that  $[r_{jk,n}]$  is a constant  $[r_{jk}]$ . Finally, and perhaps of greatest significance, the sequence of theoretical states that occurs over trials is assumed to be governed by a probability law that is specified by the theory.

Before considering the nature of the probabilities connecting the theoretical states from trial to trial, it may be noted that the existence of any probability law governing the transitions between theoretical states, along with probabilities relating theoretical states and observable responses, are sufficient to specify empirical probabilities of the kind presented in the illustration given above. The sequence of states is a stochastic process with trial outcomes  $T_{j1}, T_{j2}, \dots, T_{jn}, \dots$ . The probability of any sequence of outcomes can be calculated using the transition probabilities of the system. Given a sequence of theoretical states, the probability of any response sequence can be calculated using the probabilities relating the theoretical states and the responses. The conditional

probabilities  $p_{ik}$  have the form

$$P(M_{k,n} | S_{i,n-1}) = P(M_{k,n} | M_{k_1,1}, M_{k_2,2}, \dots, M_{k_{n-1},n-1})$$

$$\frac{P(M_{k_1,1}, \dots, M_{k_{n-1},n-1}, M_{k,n})}{P(M_{k_1,1}, \dots, M_{k_{n-1},n-1})}$$

Since the probabilities of all the sequences can be calculated, so can the conditional probabilities.

The nature of the transition probabilities of the system has a fundamental influence on the properties of the system. A desirable situation is one in which the theoretical states have the Markov property. When a theory is Markovian in its postulated states, the probability of any state  $T_{jn}$  on trial  $n$  depends only on the state of the system on trial  $n-1$  and is independent of the sequence of states that occurred on trials  $1, \dots, n-2$ . That is,

$$P(T_{jn,n} | T_{j1,1}, \dots, T_{jn-1,n-1}) = P(T_{jn,n} | T_{jn-1,n-1}).$$

The Markov property represents a kind of independence of history. In a system lacking the Markov property, the future behavior of the system is dependent both on the present state of the system and on its past behavior. If a system has the Markov property its future behavior depends only on its present state. This has far-reaching implications for the analysis of the system. If its states are Markov, then any method that permits us to specify its present state permits us to predict its future behavior, up to the uncertainty imposed by the probability laws that govern the system. If the states of a theory are not Markov, then predictions

about future behavior can be improved by obtaining information about states that occurred in the past. If this is the case, then it follows that the description of a system provided by the theory omits important distinctions. Clearly, the future behavior of the system has to depend on its present state, however it arrived there. And the finding that a theory is not Markov in its postulated states is clear evidence that the states do not give a complete description of the system.

Of course, we can assume that any description involving probabilistic relationships among states is incomplete. If a complete description were available, then the behavior of the system would be deterministic. However, the discovery of a Markovian structure provides a basis for further investigation that simplifies the problem of refining the theoretical description. If the best available theory of a process has states with the Markov property, then further investigation can be focussed on distinguishing between relevant subsets of the class of events that are grouped together in the theoretical description. Whatever one finds regarding the subsets of such events can be treated as a simple reclassification of the events. If the states of a system as described by the theory are not Markov, then variables that are relevant to the system's future behavior must be evaluated in relation to the sequence of states in the history of the system, and this will generally involve a considerable cost in theoretical complexity.

I will illustrate these ideas about Markovian theories for non-stationary systems with a theory of simple memorizing. The kind of experiment to which the theory applies involves presentation of pairs of items that are unrelated. The pairs that a subject is asked to memorize might have short words as stimuli and numbers as responses. On each

experimental trial, the experimenter presents a word and asks the subject to give the number that he thinks is correct. After the subject responds, the experimenter presents the correct answer. On the first trial, of course, the subject has to guess. However, after the subject has seen all of the pairs he can remember the correct answer on at least some of the tests, and eventually he is able to give the correct answer to all of the items.

The data from an experiment like this are analyzed in the form of sequences of responses given to the individual items. For example, on one item a subject might give the sequence of responses

0 1 0 1 1 0 0 0 ... ,

meaning that he guessed correctly on the first trial, gave an error on trial 2, a correct response on trial 3, errors on trials 4 and 5, and then correct responses from then on.

The simple model that I will describe was first developed by Bower [1]. According to the model, an individual item is learned in an all-or-none fashion. That is, at the beginning of the experiment each item is unlearned. On each study trial, there is some probability that the item becomes learned. This probability is a constant--that is, the probability of learning an item does not increase over trials during the experiment. Once an item is learned, the subject is assumed to remember it for the remainder of the experimental trials. Prior to learning an item, the subject has to guess the answer on each of the item's tests.

Putting this more formally, the theory postulates two states in which we can find an item--U and L for unlearned and learned. Each item begins in state U, and on each trial there is a constant probability called

c that the item goes from state U to state L. State L is absorbing-- that is, once an item goes into state L it stays there. This set of assumptions can be expressed in standard notation as

$$P(L_1, U_1) = (0, 1) \quad ,$$

$$[t_{j_{n-1}j_n}] = \begin{matrix} & L_n & U_n \\ \begin{matrix} L_{n-1} \\ U_{n-1} \end{matrix} & \begin{vmatrix} 1 & 0 \\ c & 1-c \end{vmatrix} \end{matrix} \quad (21)$$

where the first equation states the assumption that all the items start in state U, and the second equation states the assumption of a constant probability of a transition from U to L and the assumption that L is an absorbing state.

The final assumption links these ideas about the postulated states to probabilities of response. There are two possible responses on each trial--the subject can be correct or wrong. A correct response is denoted 0 and an error is denoted 1. While an item is in state U there is a probability, assumed to be constant, of a correct response by guessing. After the item goes into state L the probability of a correct response is assumed to be 1. This is stated in the following equation:

$$[r_{jk}] = \begin{matrix} & 0 & 1 \\ \begin{matrix} L \\ U \end{matrix} & \begin{vmatrix} 1 & 0 \\ g & 1-g \end{vmatrix} \end{matrix} \quad (22)$$

The major simplification resulting from the Markov assumption can be seen readily. As can be seen from equations (19) and (20), predictions





is needed for computation of  $I_j$  when  $j$  becomes large.) However, to provide some indication of the behavior of the statistics, I carried out calculations for higher values of  $c$  until the sums did approach asymptotic values. The results of these calculations are in Table 1. The main findings of interest in these calculations are the strong

Table 1  
Information-Theoretical Statistics for  
All-or-None Learning

Parameters	$\sum_{j=1}^{\infty} I_j(S,T)$	$\sum_{j=1}^{\infty} I_j(T,M)$	$\sum_{j=1}^{\infty} I_j(S,M)$
$c=.40, g=.25$	0.93	1.84	0.66
$c=.40, g=.50$	0.64	1.09	0.29
$c=.60, g=.25$	0.29	0.97	0.21
$c=.60, g=.50$	0.19	0.58	0.09

dependence of the amount of information transmitted on the parameter values, and the further illustrations of the fact that the dependence between empirical variables is always less strong than the dependence between either set of empirical variables and the theoretical variables.

The preceding calculations all have to do with the theoretical states considered as mediators between preceding response sequences and the response on trial  $n$ . Another view of the situation can be obtained by examining the sequence of theoretical states, without regard for the observable responses. This latter point of view is concerned with uncertainty and information transmitted at the level of the states that are postulated in the theory, and in some ways gives a more direct eval-

uation of our state of knowledge than the analysis that deals with responses, assuming the decision that the theory represents the best available understanding of the system in question.

The analyses of theoretical sequences are similar to those given earlier in connection with equations (18)-(20), except that the Markovian structure of the theory permits us to ignore states of the system occurring in the past. For any trial  $n$ , the probability of any state depends only on the state on trial  $n-1$ ; in fact, the probabilities of state-to-state transitions are constant. For trial 2, we have

$$[t_{j_1 j_2}] = \begin{array}{cc} & \begin{array}{cc} (.20) & (.80) \\ L_2 & U_2 \end{array} \\ \begin{array}{cc} (0.0) & L_1 \\ (1.0) & U_1 \end{array} & \left| \begin{array}{cc} 1 & 0 \\ .20 & .80 \end{array} \right. \end{array} \quad (27)$$

$$H_1(T) = 0.0, \quad H_2(T) = 0.50, \quad H_{12}(T \times T) = 0.50, \quad I_2(T, T) = 0.0.$$

For trial 3, the probabilities given at the top of equation (27) are the probabilities of the explanans, and the transition probabilities remain unchanged. The state probabilities on trial 3 are  $P(L_3) = .36$ ,  $P(U_3) = .64$ . The information-theoretical quantities are

$$H_3(T) = 0.65, \quad H_{23}(T \times T) = 1.08, \quad I_3(T, T) = 0.07.$$

On trial 4, the state probabilities are  $P(L_4) = .49$ ,  $P(U_4) = .51$ , leading to

$$H_4(T) = 0.69, \quad H_{34}(T \times T) = 0.97, \quad I_4(T, T) = 0.37.$$

Aside from the marked increase in simplicity, compared with the

observable sequences, the theoretical sequences also have more information transmitted. For the trials given above, the calculations given earlier for  $I(S,M)$  are 0.0, 0.04, and 0.17. The fact that values of information transmitted are higher in the theoretical sequences than in the observable sequences is not an accident. Any matrix of probabilities in the observable responses like those given in equations (18)-(20) can be derived as the products of three matrices:

$$P(S_{n-1}, M_n) = P(S_{n-1}, T_{n-1}) P(T_{n-1}, T_n) P(T_n, M_n).$$

The matrix designated by the central term is the transition matrix for the theoretical states, and we have already seen that the information transmitted by a product is no greater than the information transmitted by any of the matrices multiplied to form the product.

Both the simplicity and information-theoretical advantages of the theoretical structure are illustrated further by the analysis of information transmitted summed over trials. The analysis of this statistic for sequences of observable responses was discussed earlier; each of the calculations presented there required nearly an hour of calculation on a medium-small computer (an IBM 1800 with 16,000 words of core storage and 4 microsecond memory access). The calculations for the theoretical sequences can be done quite easily by hand. In general,

$$I_n(T,T) = H_{n-1}(T) + H_n(T) - H_{n-1,n}(T,T),$$

where

$$\begin{aligned} H_n(T) &= -P(L_n) \log P(L_n) - P(U_n) \log P(U_n) \\ &= [1-(1-c)^{n-1}] \log [1-(1-c)^{n-1}] - (1-c)^{n-1} \log (1-c)^{n-1}, \end{aligned}$$

$$\begin{aligned}
H_{n-1,n}(T,T) &= -P(L_{n-1})\log P(L_{n-1}) - P(U_{n-1},U_n)\log P(U_{n-1},U_n) \\
&\quad - P(U_{n-1},L_n)\log P(U_{n-1},L_n) \\
&= -[1-(1-c)^{n-2}]\log[1-(1-c)^{n-2}] - (1-c)^{n-1}\log(1-c)^{n-1} \\
&\quad - c(1-c)^{n-2}\log c(1-c)^{n-2} .
\end{aligned}$$

Combining terms and summing across values of  $n$ ,

$$\sum_{j=2}^{\infty} I_j(T,T) = \log c - \frac{(1-c)}{c} \log(1-c) - \sum_{k=1}^{\infty} [1-(1-c)^k] \log[1-(1-c)^k]$$

For values of  $c$  of .20, .40, and .60 the sums  $\sum I_n$  for the theoretical sequences are 5.67, 3.99, and 1.16. The last two values can be compared with values given in Table 1 for the observable sequences.

Since there is a simpler structure and greater information transmitted in the theoretical sequence than there is in the sequence of observations, the theory provides an advantageous basis for developing new knowledge. The development of new measurement techniques for increasing dependencies between observations and the theoretical states, and the refinement of theory to provide increased information transmitted in the trial-to-trial transitions both constitute additions to knowledge that are frequent in scientific investigation, and that are explicable within the framework of the present analysis.

### Conclusions

In presenting the information-theoretical analysis of theoretical entities of this paper, a number of rather routine calculations have been carried out. To some extent, the significance of these analyses is just that they can be carried out. The analyses of this paper serve as an

existence proof that it is possible to perform analyses that are relevant to evaluating our state of knowledge when we have statistical knowledge about a system. It is universally recognized that theoretical entities can serve as an heuristic aid in development of new empirical knowledge about a system, and in some cases can provide a substantial simplification in the representation of knowledge. These facts are clarified and specified to some extent by the present analysis. We have seen that the information transmitted by dependencies between a set of empirical variables and a set of theoretical variables is generally greater than the information transmitted between sets of empirical variables, and this fact clarifies the usefulness of postulated entities in guiding empirical investigations. The role of theoretical entities in relation to simplicity is especially striking in nonstationary systems, particularly when the theoretical system has the Markov property.

## References

- [1] Bower, G. H., "Application of a model to paired-associate learning," Psychometrika, vol. 26, 1961, pp. 255-280.
- [2] Greeno, J. G. "Evaluation of Statistical Hypotheses Using Information Transmitted," Philosophy of Science, vol. 37, 1970, in press.
- [3] Krantz, D. H., "Threshold Theories of Signal Detection," Psychological Review, vol. 76, 1969, pp. 308-324.
- [4] Luce, R. D., "A Threshold Theory for Simple Detection Experiments," Psychological Review, vol. 76, 1969, pp. 308-324.

## DOCUMENT CONTROL DATA - R &amp; D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

ORIGINATING ACTIVITY (Corporate author) University of Michigan, Human Performance Center Department of Psychology, Ann Arbor, Michigan	20. REPORT SECURITY CLASSIFICATION <b>UNCLASSIFIED</b>
25. GROUP	

REPORT TITLE  
**THEORETICAL ENTITIES IN STATISTICAL EXPLANATION (TALK BASED ON THIS PAPER GIVEN AT MEETINGS OF THE PHILOSOPHY OF SCIENCE ASSOCIATION, BOSTON, OCTOBER 1970)**

DESCRIPTIVE NOTES (Type of report and inclusive dates)  
**Scientific Interim**

AUTHOR(S) (First name, middle initial, last name)

**James G. Greeno**

REPORT DATE <b>October, 1970</b>	70. TOTAL NO. OF PAGES <b>35</b>	75. NO. OF PAGES <b>4</b>
CONTRACT OR GRANT NO. <b>AF49(638)1736 ARPA</b>	90. ORIGINATOR'S REPORT NUMBER(S) <b>Memorandum Report No. 12</b>	
PROJECT NO. <b>AO 461</b>	95. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) <b>AFOSR-TR-71-0863</b>	
<b>61101D</b>		
<b>681313</b>		

DISTRIBUTION STATEMENT

**This document has been approved for public release and sale; its distribution is unlimited.**

SUPPLEMENTARY NOTES <b>TECH, OTHER</b>	12. SPONSORING MILITARY ACTIVITY <b>AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (NL) 1400 WILSON BLVD ARLINGTON, VIRGINIA 22209</b>
---	--

ABSTRACT

An information-theoretical analysis of theories having statistical hypotheses and postulated entities is carried out. The analyses show that it is possible to perform analyses that are relevant to evaluating our state of knowledge when we have statistical knowledge about a system. It is universally recognized that theoretical entities can serve as an heuristic aid in development of new empirical knowledge about a system, and in some cases can provide a substantial simplification in the representation of knowledge. These facts are clarified and specified to some extent by the present analysis. We have seen that the information transmitted by dependencies between a set of empirical variables and a set of theoretical variables is generally greater than the information transmitted between sets of empirical variables, and this fact clarifies the usefulness of postulated entities in guiding empirical investigations. The role of theoretical entities in relation to simplicity is especially striking in nonstationary systems, particularly when the theoretical system has the Markov property.

FORM 1 NOV 65 473

**UNCLASSIFIED**

Security Classification